Project 1: Fundamental frequency of vowels

MATH-516 Applied Statistics

Linda Mhalla

2025-02-17

1 Data: Acoustic analysis of vowels

- Data on acoustic measurements of vowels produced by American English speakers
  - data from Hillebrand et al. (1995)
  - measurements on a group of children, men, and women were recorded
  - the fundamental frequency f_0 of a vowel's sound wave was measured (in Hz)
- Only binned data are available (and the grid is not equidistant)
  - available data: the number of recorded vowels (variable `count`) and the fraction (variable `percentage`) of vowels' fundamental frequency belonging to each bin (given by `start_point` and `end_point`)

	X	start_point	end_point	count	percentage
1	1	90	107	67	4.016787
2	2	107	115	76	4.556355
3	3	115	121	61	3.657074
4	4	121	130	73	4.376499
5	5	130	136	60	3.597122
6	6	136	144	64	3.836930

2 The Goal

Simulate the vowels' frequency f_0 from a distribution, which is as close as possible to the observed data, in order to study the different characteristics of these sounds in the time domain

- i.e., the goal is to do Monte Carlo: **how to simulate vowels' frequencies that are compatible with the data?**

Expert knowledge: a mixture of two log-normal distributions is a good model for the lowest frequency f_0

3 Tasks for You

1. Is the assumption viable, i.e., is bi-log-normal distribution a reasonable model for the data?
   - simple exploration of the data
2. Fit the bi-log-normal distribution in order to be able to simulate the data easily using
   - jittering and EM algorithm OR direct optimization (e.g., local search starting from the jittered EM result), AND
   - a Bayesian approach
3. Test whether the fundamental frequencies come from a bi-log-normal distribution
   - parametric bootstrap and goodness of fit

4 MATH-517 Content

- Week 1: Introduction & Software & Data Considerations
- Week 2: Graphics & Visualization
- **Week 3: Kernel Density Estimation**
- Week 4: Non-parametric Regression
- **Week 5: Cross-validation**
- **Week 6: EM Algorithm**
- **Week 7: EM Algorithm**
- **Week 8: Monte Carlo**
- **Week 9: Bootstrap**
- **Week 10: Bootstrap**
- **Week 11: Bayesian Computations**
- **Week 12: Bayesian Computations**
- Week 13: Decision Trees
- Week 14: $\emptyset$
  - Weeks in bold are pertinent to Project 1
  - Weeks 1-2 established the workflow needed for all the projects